

<https://helda.helsinki.fi>

Molecular ecology of the yet uncultured bacterial Ct85-cluster in the mammalian gut

Hynönen, Ulla

2020-04

Hynönen , U , Zoetendal , E G , Virtala , A-M K , Shetty , S , Hasan , S , Jakava-Viljanen , M , de Vos , W M & Palva , A 2020 , ' Molecular ecology of the yet uncultured bacterial Ct85-cluster in the mammalian gut ' , Anaerobe , vol. 62 , 102104 . <https://doi.org/10.1016/j.anaerobe.2019.102104>

<http://hdl.handle.net/10138/319587>

<https://doi.org/10.1016/j.anaerobe.2019.102104>

cc_by_nc_nd

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

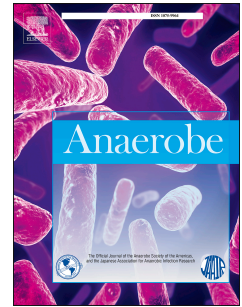
This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Journal Pre-proof

Molecular ecology of the yet uncultured bacterial Ct85-cluster in the mammalian gut

Ulla Hynönen, Erwin G. Zoetendal, Anna-Maija K. Virtala, Sudarshan Shetty, Shah Hasan, Miia Jakava-Viljanen, Willem M. de Vos, Airi Palva



PII: S1075-9964(19)30176-3

DOI: <https://doi.org/10.1016/j.anaerobe.2019.102104>

Reference: YANAE 102104

To appear in: *Anaerobe*

Received Date: 17 May 2019

Revised Date: 11 September 2019

Accepted Date: 23 September 2019

Please cite this article as: Hynönen U, Zoetendal EG, Virtala A-MK, Shetty S, Hasan S, Jakava-Viljanen M, de Vos WM, Palva A, Molecular ecology of the yet uncultured bacterial Ct85-cluster in the mammalian gut, *Anaerobe* (2019), doi: <https://doi.org/10.1016/j.anaerobe.2019.102104>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Ltd.

Molecular ecology of the yet uncultured bacterial Ct85-cluster in the mammalian gut

**Ulla Hynönen^{a*}, Erwin G. Zoetendal^b, Anna-Maija K. Virtala^a, Sudarshan Shetty^b, Shah Hasan^{c1},
Miia Jakava-Viljanen^{a2}, Willem M. de Vos^{b,d}, Airi Palva^a**

^aDepartment of Veterinary Biosciences, Veterinary Microbiology and Epidemiology, 66 PB, 00014
University of Helsinki, Finland

^bLaboratory of Microbiology, Wageningen University & Research, 6708 PB Wageningen, The
Netherlands

^cDepartment of Production Animal Medicine, University of Helsinki, Paroninkuja 20, 04920
Saarentaus, Finland

^dHuman Microbiome Research Program, Faculty of Medicine, University of Helsinki, Finland

*Corresponding author

¹Present address: Hankkija Oy, 05800 Hyvinkää, Finland

²Present address: Finnish Food Authority, Mustialankatu 3, 00790 Helsinki, Finland

Email addresses:

Ulla Hynönen: ulla.hynonen@helsinki.fi

Erwin Zoetendal: erwin.zoetendal@wur.nl

- 21 Anna-Maija Virtala: anna-maija.virtala@helsinki.fi
- 22 Sudarshan A. Shetty: sudarshan.shetty@wur.nl
- 23 Shah Hasan: shah.hasan@helsinki.fi
- 24 Miia Jakava-Viljanen: miiajakavaviljanen@gmail.com
- 25 Willem de Vos: willem.devos@wur.nl
- 26 Airi Palva: airi.palva@helsinki.fi

Abstract

In our previous studies on irritable bowel syndrome (IBS) –associated microbiota by molecular methods, we demonstrated that a particular 16S rRNA gene amplicon was more abundant in the feces of healthy subjects or mixed type IBS (IBS-M) –sufferers than in the feces of individuals with diarrhea-type IBS (IBS-D). In the current study, we demonstrated that this, so called Ct85-amplicon, consists of a cluster of very heterogeneous 16S rRNA gene sequences, and defined six 16S rRNA gene types, a to f, within this cluster, each representing a novel species-, genus- or family level taxon. We then designed specific PCR primers for these sequence types, mapped the distribution of the Ct85-cluster sequences and that of the newly defined sequence types in several animal species and compared the sequence types present in the feces of healthy individuals and IBS sufferers using two IBS study cohorts, Finnish and Dutch. Various Ct85-cluster sequence types were detected in the fecal samples of several companion and production animal species with remarkably differing prevalences and abundances. The Ct85 sequence type composition of swine closely resembled that of humans. One of the five types (d) shared between humans and swine was not present in any other animals tested, while one sequence type (b) was found only in human samples. In both IBS study cohorts, one type (e) was more prevalent in healthy individuals than in the IBS-M group. By revealing various sequence types in the widespread Ct85-cluster and their distribution, the results improve our understanding of these uncultured bacteria, which is essential for future efforts to cultivate representatives of the Ct85-cluster and reveal their roles in IBS.

Keywords: uncultured; intestinal microbiota; IBS; 16S rRNA; typing

Introduction

Irritable bowel syndrome (IBS) is a functional bowel disorder affecting on average 11% of the population worldwide, with significant geographic variation in prevalence[1]. It is characterized by intestinal discomfort and an altered bowel habit, based on which diarrheal (IBS-D), constipation (IBS-C) and mixed (IBS-M) types of IBS can be distinguished[2]. The etiology of IBS is multifactorial and different between individuals; intestinal barrier function, immune and enteroendocrine functions as well as nervous signaling apparently have a role in IBS development[3]. In addition, the role of gut microbiota as one of the contributors of IBS is generally recognized[3-5].

We have previously compared by 16S rRNA gene-based methods the intestinal microbiotas of healthy individuals and of those suffering from different types of IBS in Finland, and found 16S rRNA gene amplicons, detectable by specific primers, that are either enriched or depleted in healthy individuals or in IBS sufferers[6-8]. Recently, we also isolated in pure culture and characterized one of the IBS-D –associated phylotypes, RT94[9], which was simultaneously isolated and named *Sellimonas intestinalis* by Seo et al[10]. As a continuation of these studies, we next chose to isolate and characterize bacteria belonging to the 16S rRNA gene group that was amplified by so called Ct85-primers[8]. This sequence group was originally [8] supposed to comprise a coherent phylotype and it was named Ct85 according to its similarity (%) to the 16S rRNA gene sequence of the phylogenetically closest cultured species at that time, *Clostridium thermosuccinogenes*. Previously we have shown in a Finnish study that Ct85 16S rRNA gene copy numbers are elevated in healthy individuals and IBS-M sufferers compared to IBS-D sufferers[8], and that these numbers in individuals with IBS rise to the levels seen in healthy persons during a probiotic intervention, while remaining low in the placebo group[11]. Concomitant with the rise in Ct85-amplicon abundance, IBS symptom scores in the probiotic group are lowered[12]. The cultivation and characterization of a bacterium representing this 16S rRNA cluster and revealing its

interactions might thus provide invaluable information about the ecology of dysbiotic and non-dysbiotic microbiota as well as knowledge aiding in IBS diagnostics and subject stratification.

Our extensive, systematic efforts to isolate in pure culture a representative of the Ct85 bacterial cluster were however unsuccessful. To get insight into the diversity of bacteria belonging to this cluster, we, in this study, sequenced 21 partial, Ct85-cluster-representing 16S rRNA genes in our previously constructed clone libraries[7, 8] to near full length and aligned these sequences with the three completed Ct85-cluster 16S rRNA gene sequences we have previously deposited in public databases. Three groups of sequences with sequence identities of 97.2-100% in each, and one group with 95.3-99.9% internal sequence identity, could be distinguished in this sequence panel. PCR primers specific for these four groups, and primers specific for two 16S rRNA gene types represented by a single sequence in the panel were designed and checked for specificity against 16S rRNA gene sequences in public databases. These type-specific primers were then used to survey the presence and abundance of the different Ct85-cluster sequence types in various companion and production animal species. In addition, the fecal DNA samples of two study cohorts of IBS sufferers and healthy controls in two countries were typed with the new primers with intention to reveal anticipated correlations between the prevalence or abundance of a particular Ct85-cluster sequence type and the healthiness or IBS-type of the individuals. In summary, in this work we aimed at obtaining necessary basic information improving our understanding of this large group of uncultured bacteria and thus forming the basis for future efforts to isolate representatives of the Ct85-cluster in pure culture, a goal crucial to understanding the ecology and role of these bacteria in intestinal health and IBS.

Materials and Methods

Sequencing, sequence alignment, primer design and phylogenetic analysis

16S rRNA gene sequencing by the Sanger method was performed in the Institute of Biotechnology, University of Helsinki, and the near full-length sequences were published in the European Nucleotide Archive (ENA) under the accession numbers LR595726-LR595746 (study PRJEB31919). The 16S rRNA gene alignment and the percent identity matrix were constructed by Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Primers specific for the Ct85-cluster sequence types were designed based on the alignment, checked for physicochemical parameters using OligoAnalyzer (<https://eu.idtdna.com/pages/tools>), NetPrimer (<http://www.premierbiosoft.com/netprimer/index.html>) and Multiple Primer Analyzer (<https://www.thermofisher.com/fi/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>) and tested for specificity *in silico* using the “Primer Pair Specificity Checking” function of PrimerBlast (database “nr”, organism “bacteria, taxid:2”, stringent primer specificity parameters, i.e. no mismatches allowed) (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). All the near full-length hits obtained this way were aligned by pairwise blastn with the correct target sequences (i.e. representative 16S rRNA gene sequences of each sequence type in our panel), and primer pairs producing hits with 96-100% similarity with the correct target sequences were accepted. In case of primer pairs c and d, however, the primer pairs were accepted although occasional hits with 88-93% overall similarity, as determined by Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) were obtained. All the primers in this work were manufactured by ImmunoDiagnostic Oy (Hämeenlinna, Finland). The phylogenetic tree was constructed in MEGA7 [13] using the Maximum Likelihood method based on the Tamura-Nei model [14]. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the

topology with superior log likelihood value. The tree with the highest log likelihood (-7801,16) is shown. Codon positions included were 1st, 2nd, 3rd and noncoding. All positions containing gaps and missing data were eliminated.

Human subjects

Fecal DNA samples of two IBS study cohorts were used in this work; one of the studies was performed in Finland[6] and another in the Netherlands (manuscript in preparation). Rome II or Rome III diagnostic criteria for IBS were used in the Finnish and Dutch study, respectively, for subject selection, and both studies compared the microbiotas of healthy individuals and symptom-typed IBS sufferers at different time points. Time point zero samples of the latter study were used; in case of the former study, 3 month samples were occasionally used if zero samples were not available. The fecal samples of the Finnish study have also been investigated in several previous studies[7, 8, 12, 15-19]. The numbers of fecal samples in the different subject groups of both cohorts are shown in Table 1.

In both the Finnish and Dutch cohort, the informed consent of the participants was obtained and the studies were performed in compliance with local laws and institutional guidelines and approved by local ethical committees, i.e. the Joint authority for the hospital district of Helsinki and Uusimaa and the ethical committee of VTT Technical Research Center of Finland (Finnish study), and Wageningen University & Research (Dutch study).

Animals used for fecal sampling

The details of the sows and piglets and their housing are described in [20]. The wild boars were reared on a single outdoor farm in southern Finland and they were grazing and rooting for their food; in addition, their diet was subsidized with different grains, seasonal fruit and vegetables. Pet dogs and part of the cats were privately owned and living in different environments (urban/rural);

part of the cats were housed in a rehoming center for lost animals. The laboratory rodents, an NMRI mouse and a Wistar rat, were housed in the Laboratory Animal Center of the University of Helsinki and fed common commercial rodent feed. The cows were all living on the Viikki Research Farm (Helsinki, Finland) in the same cowshed and eating the same commercial feed indoors and grass outdoors.

No approval by an ethical committee was required for the animal studies, as only rectal or extruded fecal samples were used, requiring no invasive procedures.

Fecal DNA isolation

The fecal samples were kept at +4 °C and processed on the same day as collected, or immediately frozen at -20 °C after collection for later use. Bacterial DNA was isolated from all the samples by the repeated bead beating method as described by Salonen et al [21], with the exception of wild boar fecal samples, from which the DNA was isolated as described in [22]. The sow and piglet fecal DNA samples were from the work of Hasan et al[20].

Quantification of Ct85-cluster 16S rRNA genes

The abundance of Ct85-cluster 16S rRNA genes in human or animal fecal DNA samples was determined by qPCR with primers described by Lyra et al[8] with the near full-length 16S rRNA gene AM275406.2, amplified from the library clone AP07K.11[7] with primers S-D-Bact-0043-a-S-22 (5' CGCACATAAGACATGCAAGTCG 3', F) and S-D-Bact-1521-a-A-20 (5' CCTTCCGATACGGCTACCTT 3', R), as a standard. The reaction mixtures consisted of SYBR Select Master Mix (ThermoFisher Scientific), 400 nM of Ct85-specific primers[8] and 100 ng of fecal DNA as templates in a 20 µl reaction volume. A cycling mode with an annealing step at 57 °C for 15 s and extension at 72 °C for 1 min and 40 cycles was used, according to the Sybr Select Master Mix manufacturer's instructions. qPCR reactions with universal primers[23] were performed similarly

but using 1 ng of fecal DNA as templates. All the qPCR reactions in this work were performed with the Stratagene MX3000P qPCR instrument (AH Diagnostics, Helsinki, Finland).

Determination of the presence and relative abundance of Ct85-cluster sequence types in fecal samples

To increase detection sensitivity when typing Ct85-cluster 16S rRNA gene sequences, all 16S rRNA genes were first amplified from fecal DNA with universal bacterial 16S rRNA gene primers (PCR I), after which the reaction mixtures were diluted and used as templates in end-point PCR reactions with sequence type-specific primers (PCR II). For determining the relative abundance of each type, the PCR I reaction mixtures positive for a particular primer pair were then used as templates in qPCR reactions with the same type-specific primers and with universal bacterial 16S rRNA qPCR gene primers (quantitative PCR II).

PCR I

16S rRNA genes were amplified using of 0.5 μ M of universal bacterial 16S rRNA gene primers[24], 50 ng of template, 0.25 U of Phusion High Fidelity Polymerase (ThermoFisher Scientific) and 200 μ M of each dNTP (ThermoFisher Scientific) in 1 x HF buffer in a 12.5 μ l reaction volume with cycling conditions recommended by the enzyme manufacturer (T_{ann} 52 °C).

PCR II

For type-specific end-point PCR reactions (PCR II), the amplification conditions for each primer pair (annealing temperatures and template-, primer- and DNA-polymerase concentrations) were first optimized using plasmids in previously constructed gene libraries[7], carrying 16S rRNA genes of the corresponding sequence types, as templates. 16S rRNA genes of all the other sequence types were used as negative controls. The primer sequences and optimized cycling conditions for end-

point PCR II reactions are detailed in Table 2. In the optimized conditions, 0.5 μ M of type-specific primers were used, and 0.025 μ l of the PCR I reaction mixture (based on testing with different amounts of template), or 5 ng of plasmid template in positive controls, were used as templates with 0.25 U of Phusion High Fidelity Polymerase (ThermoFisher Scientific) and 200 μ M of each dNTP (ThermoFisher Scientific) in 1 x HF buffer in a 12.5 μ l reaction volume.

Quantitative PCR II

The conditions for type-specific qPCR reactions (qPCR II) were first optimized using near full-length 16S rRNA genes of the different Ct85 types as templates; these genes were amplified from the 16S rRNA gene library plasmids[7] with primers S-D-Bact-0043-a-S-22 and S-D-Bact-1521-a-A-20 (see “Quantification of Ct85-cluster 16S rRNA genes” above for primer sequences). The templates were used as ten-fold dilution series from 10^2 to 10^7 copies/well, and different cycling programs, annealing temperatures and primer concentrations were tested for each template with SYBR Select Master Mix (ThermoFisher Scientific) as the base for the amplification reactions, according to the manufacturer’s instructions. As negative controls, 16S rRNA gene amplicons from selected 16S rRNA gene types (a-f) were used, i.e. those that in the corresponding end-point PCR with the same primers had at some point shown some unspecific products.

The qPCR II reactions were then performed under the optimized conditions using the 16S rRNA gene amplicons described above as standards: a cycling mode with a combined annealing-extension step at 60 °C for 1 min with 40 cycles was applied, and 400 nM of type-specific primers and 0.05 μ l of unpurified 16S rRNA gene amplification (PCR I) products (volume based on prior testing) were used as templates in a 20 μ l reaction volume.

The total numbers of 16S rRNA genes in the PCR I reaction mixtures were determined by qPCR with the near full-length 16S rRNA gene from *Lactobacillus amylovorus*, amplified from its genomic

211 DNA with universal bacterial 16S rRNA gene primers[24], as a standard. In these reaction mixtures
 212 with SybrSelect Master Mix, 400 nM of universal bacterial 16S rRNA gene qPCR primers[23] were
 213 used, and 5×10^{-6} μ l of the unpurified PCR I amplification products (based on prior testing) were
 214 used as templates in a 20 μ l reaction volume. The cycling was performed with an annealing step at
 215 57 °C for 15 s, extension at 72 °C for 1 min and 40 cycles. In both the optimization stage- and final
 216 qPCR assays with any primers, three replicates of standards, samples and controls were used, and
 217 a “no template”-control was present in every qPCR run of this study.

218 **Survey of Ct85 sequence types in public sequence databases**

219 To investigate the ecological occupancy of the newly defined sequence types of the Ct85-cluster,
 220 we searched the publicly available datasets via the MetaMeta-database available at
 221 <http://mmdb.aori.u-tokyo.ac.jp/index.html> [25] using default settings. Particular sequence types
 222 were also searched in the IMNGS-database available at <https://www.imngs.org/> [26]. In addition,
 223 we leveraged the recently created databases of 16S rRNA genes linked to recently reported
 224 metagenome assembled species-level genome bins (SGB) from the human gut [27]. All the 16S
 225 rRNA gene sequences of SGBs reported by [27] were merged to create a custom blast nucleotide
 226 database. A blastn (megablast) search was performed with the following parameters: *task*,
 227 megablast; *e-value*, 0.001 and *max_target_seqs*, 10[28]. PrimerBlast searches against public
 228 nucleotide databases were performed as described under “Sequencing, sequence alignment,
 229 primer design and phylogenetic analysis” above.

230 **Statistical analysis**

231 Human data were analyzed using IBM SPSS for Windows (IBM Corp; Version 25.0, Armonk, NY).
 232 The association between the four subject groups (healthy, IBS-D, IBS-C and IBS-M) and any of the
 233 six Ct85 sequence types (a to f) was studied using Fisher’s exact test. To adjust for multiple

comparisons (several sequence types separately with patient groups), the Benjamini-Hochberg method was applied using the FDR (false discovery rate) calculator (<https://tools.carbocation.com/FDR>). Statistical significance was considered when $P \leq 0.05$. The association between the number of sequence types (0 to 6) was studied using independent samples Kruskal-Wallis test with FDR correction. The 95% confidence intervals for all percentages were calculated by Wilson's method[29] using an internet-based epidemiological calculator (<http://epitools.ausvet.com.au/content.php?page=CIPProportion>).

Results

Definition of Ct85 sequence types

The pairwise similarities of the 24 Ct85-cluster 16S rRNA gene sequences cloned by Kassinen et al[7] and now sequenced to near full-length, together with the 16S rRNA sequences of their closest cultured relatives, are shown in Figure 1A. Only one of the 24 sequences (from clone AP10.R.212) originated from the library made of IBS-D sufferers' fecal DNA, while most of the sequences were from the libraries made of healthy controls' or IBS-M sufferers' fecal DNA[7]. Two 16S rRNA gene sequences from the healthy individuals' clone library (clones AP07K.333 and AP07K.377) differed remarkably from all the other 16S rRNA gene sequences in the panel. Indeed, the fully sequenced 16S rRNA gene of the clone AP07K.333 was found by pairwise blastn to have 95.1% identity to *Roseburia faecis* 16S rRNA gene (NR_042832.1, cover 90%), and the sequence of the clone AP07K.377 94.3% sequence identity to the 16S rRNA gene of *Ruminococcus bromii* (NR_025930.1, cover 87%). The corresponding identity values by Clustal Omega were 92.3% for both, indicating that bacteria carrying these 16S rRNA genes probably belong to the families *Lachnospiraceae* and *Ruminococcaceae*, respectively. These two diverging 16S rRNA gene sequences, resembling those of cultured bacteria, were excluded from further analysis. All the

257 sequences in the remaining set had only approximately 86% identity by pairwise blastn (and even
 258 less by Clustal Omega) to the 16S rRNA genes of closest known cultured species
 259 (*Hungateiclostridium aldrichii*, NR_026099.2, *H. sellulolyticum*, NR_025917.1, *Pseudoclostridium*
 260 *thermosuccinogenes*, NR_119284.1 or *Ruminiclostridium hungatei*, NR_117165.1). In this
 261 remaining set of 22 sequences, four groups of 2-7 sequences in each, with internal similarities of
 262 95.3-100% could be distinguished (framed in Figure 1A). PCR primers specific for these four
 263 groups, as well as for the single sequences from clones AP10R.212 (from the IBS-D library) and
 264 AP07S.345, were designed and used for typing the Ct85 16S rRNA gene sequences in human and
 265 animal fecal DNA samples. Figure 1B shows the phylogenetic relationships between the newly
 266 defined sequence groups, referred to as “Ct85 sequence types” below, and their closest cultured
 267 relatives. Based on the sequence similarities of our 16S rRNA gene sequences to each other and to
 268 those (of uncultured bacteria) available in public databases, the newly defined sequence types
 269 would tentatively correspond to species (type b), genera (types a, e and f) and families (types c
 270 and d).

271 **Ct85-cluster representatives and Ct85 sequence types in humans**

272 We first determined the abundance of all Ct85-cluster 16S rRNA gene sequences in the fecal
 273 samples of our two study cohorts by qPCR. Numbers of the fecal samples in the two cohorts are
 274 shown in Table 1. All the 50 Finnish samples and 97% of the 100 Dutch samples were positive
 275 when tested with Ct85-targeting primers. As a part of the 16S rRNA gene amplifications (PCR I
 276 reactions) of the Dutch samples failed, the final, typed Dutch sample set consisted of 44 samples
 277 from healthy controls and 25 samples from IBS sufferers (Table 1). In the sequence-typed samples,
 278 the two cohorts combined, the Ct85-amplicon was detectable at different levels ranging from the
 279 hardly detectable 3.6×10^3 copies / g of feces to 2.2×10^{10} copies / g of feces, corresponding to

proportions of $4 \times 10^{-9} - 1 \times 10^{-2}$ of total 16S rRNA genes, with no remarkable differences in abundances between the two cohorts (data not shown).

The prevalences (%) of the six Ct85 sequence types (a-f) among healthy controls and among individuals of the different IBS symptom groups are shown in Figure 2. In the Finnish cohort (Figure 2A), sequence type c was significantly more prevalent in healthy individuals' than in IBS-D sufferers' samples ($P = 0.039$), and type b almost significantly more prevalent in individuals with IBS-M than in those with IBS-D ($P = 0.052$); these results are in accordance with our previous data showing differences in Ct85-amplicon abundances between these subject groups in the same Finnish sample material [8]. In addition, types d and f were in this study shown to be significantly more prevalent in the IBS-C group than in the IBS-D group samples ($P = 0.039$ and $P = 0.024$, respectively). Furthermore, type e was lacking completely from the IBS-M group, but was rather prevalent (70%) in healthy individuals' samples, making a significant difference between these groups ($P = 0.024$). In the Dutch cohort (Figure 2B), however, the difference in prevalence of type c between healthy individuals and IBS-D sufferers, or the differences in the prevalences of types d and f between IBS-C and IBS-D groups were not seen. Moreover, the difference in the prevalence of type b between the IBS-D and IBS-M groups could not be evaluated, as this type was not present in any Dutch IBS-D or IBS-M sample. However, the very rare occurrence of type e in IBS-M sufferers' samples compared to those of healthy individuals, observed in the Finnish cohort, was evident also in the Dutch cohort, although the difference in the latter did not reach statistical difference. When combining the Finnish and Dutch datasets, a statistically significant difference ($P = 0.006$) in the occurrence of sequence type e was seen between these two subject groups, while combining the datasets did not reveal any further significant differences in sequence type occurrences between groups.

The relative abundances of the Ct85 sequence types a to e in IBS sufferers' and healthy controls' fecal DNA samples are shown in Figure 3. The low incidence of most sequence types in the IBS groups hampered comparisons between subject groups, and the differences in prevalences described above could not be recognized in abundances. Notably, however, sequence type d was very low in abundance in the only IBS-D sample in which it was present (Dutch cohort). It was generally of rather low prevalence in both cohorts, but especially rare or even absent in IBS-D sufferers' samples. It was hardly detectable in the single Dutch IBS-D sample by qPCR (proportion 2.9×10^{-10} of total 16S rRNA genes), while the relative abundance of this type in healthy controls' samples in the same cohort was in the range of $9.7 \times 10^{-10} - 4.4 \times 10^{-5}$, and in the single IBS-C sufferers' sample even reasonably high (3.8×10^{-3}).

We also found that the numbers of different Ct85 sequence types per individual were higher in healthy controls than in IBS-D sufferers (Supplementary Figure 1). This difference was statistically significant for the Finnish cohort ($P = 0.03$, Kruskal-Wallis-test), while in the Dutch cohort, this seemingly evident difference did not reach statistical significance ($P = 0.51$), probably due to the low number of IBS-D samples typed. As shown in Supplementary Figure 1, in the Finnish cohort, the number of Ct85 sequence types most commonly found among healthy controls was three or four (panel A2), while IBS-D sufferers did not have any or had only one or two recognizable types (panel A4). In the Dutch cohort the most common number of sequence types was three in healthy controls (panel B2) and one in IBS-D sufferers (panel B4). The trend of healthy controls having higher numbers of sequence types was clear for both cohorts also when comparing healthy controls to IBS sufferers as a group (panels A2 vs A3 and B2 vs B3).

Representatives of the Ct85-cluster in companion and production animals

Screening of public databases revealed that 16S rRNA gene sequences closely related to those of the Ct85-cluster were not only detected in human intestinal samples, but also in samples from the pig intestinal tract. Hence, we determined the abundances of Ct85-cluster 16S rRNA genes in fecal samples collected from reared wild boars as well as from domestic pigs and their piglets. To get a glimpse if representatives of the Ct85-cluster can also be found in other domestic animals, we randomly selected feces from 19 dogs, 17 cats, and 12 cows and tested these samples for the presence of Ct85-cluster 16S rRNA genes by qPCR (Table 3). All the 25 sows tested had rather high numbers of Ct85-cluster 16S rRNA genes in their feces. Their seven piglets tested were also clearly positive for Ct85-cluster members, and Ct85-cluster 16S rRNA genes were detected also in the feces of all the tested reared wild boars, albeit at somewhat lower numbers. In dogs and cats the variation in the prevalence of Ct85-cluster 16S rRNA genes was higher. Twelve of the 19 dogs tested (63%) had detectable numbers of Ct85-cluster 16S rRNA genes in their feces, but the copy numbers were generally low, with only one dog having 5×10^7 16S rRNA gene copies /g in its feces. In cats, 14 out of 17 (82%) had Ct85-cluster members in the feces, but the inter-individual variation in copy numbers was high (7.5×10^2 - 7.5×10^7 /g). Ct85-cluster representatives were rather abundant also in the feces of all the 12 cows tested.

Ct85 sequence types in companion and production animals

Due to the failure of PCR I of some animal samples, we were able to determine the prevalence of Ct85 sequence types in the fecal samples of 14 sows, six of their piglets, two reared wild boars, 11 dogs and 12 cats (Table 4). The two most common sequence types in the fecal samples of healthy humans, e and f (see Figure 2), were also among the most common types in the pool of animals fecal samples tested, as they were found in at least one sample of each animal species tested in this study. However, animal species differed remarkably in their sequence type compositions. The

Ct85 sequence type composition most similar to that of humans was found in adult swine. In swine fecal samples, all the same Ct85 types were detected as in humans, excluding type b, not present in any animal sample. As in healthy humans, sequence type f was the most common one in swine, and also type e was among the three most often detected ones in both species. The rest of the sequence types differed in prevalence between humans and swine, type a and d being more prevalent in adult swine than in humans. The common swine sequence type d was also detected in every one-week old, suckling piglet tested. The presence of Ct85 sequence types in the fecal samples of the six sow-piglet pairs tested is shown in Supplementary Table 1. The Ct85 sequence type composition of wild boars resembled that of swine.

In contrast, dogs and cats had clearly different sequence type profiles and prevalences. In line with the generally low incidence and very low numbers of Ct85-cluster members in dog fecal samples, our type-specific primers detected Ct85-cluster members in the fecal samples of only two Ct85-cluster positive dogs out of the 11 tested; these samples represented only the sequence types a, e and f. In the other dogs positive for Ct85-primers, either the numbers of Ct85-cluster representatives have been too low to be detected by the type-specific primers or these dogs have had Ct85-cluster types not detectable by them. The same sequence types as in dogs were also detected in the fecal samples of cats.

The relative abundances of Ct85 sequence types a-e in the fecal samples of animals are shown as blue-rimmed dots in Figure 3. No striking differences were observed between healthy humans and swine in the abundances of the types a, c, d and e, shared by these two species. For example, type c, rarely detected in swine, was nonetheless of equal abundance as in humans, when present. Likewise, in both humans and swine, type e was usually of low abundance. Interestingly, sequence

type e was the only frequently detected one in cats and most cats had higher numbers of type e in their feces than did swine.

Ct85 sequence types in public sequence databases

To investigate the ecological occupancy of the newly defined sequence types of the Ct85-cluster, we utilized the MetaMetaDB-search [43], a tool to search a 16S rRNA gene database containing genes from 61 different habitats including 13 animals. The results of the MetaMetaDB-search are presented in Supplementary Figure 2. Based on this database, bacteria resembling Ct85 sequence types a, c and d (at the 97% minimum identity level) are common inhabitants of both the pig intestine and the human gut, but are not found in other animals or i.e. in the oral, lung, marine or soil environments, which supports our findings about the occurrence of these sequence types. In addition to human and pig gut, type d-like sequences in this search were found in bioreactor sludge/anaerobic digesters, bioreactors and freshwater, with no clear preference for the human gut over the other environments, which was also in line with our finding that type d sequences were rather uncommon in humans. A PrimerBlast search with type d primers further revealed 16S rRNA gene sequences closely resembling those of type d (98-100% identity) to be present in human, swine and orangutan stool as well as in an Egyptian lake sediment.

Type e-like sequences at the 99% identity level, in turn, were commonly found in both human and pig intestine by using the IMNGS-search platform [44], which also utilizes very large 16S rRNA gene amplicon datasets, and type f-like sequences were found to be present in human [30-36] and pig [37] fecal samples by blastn or PrimerBlast searches. The generally low relative abundance of type e sequences in animal fecal DNA, as observed in the current study, may explain why sequences of this type have not been detected in metagenomic studies represented by the MetaMetaDB-search. Bacteria representing sequence type e, detected in every animal species tested in this

study and as the predominant sequence type in cats, may also be present in preadolescent turkey ceca; these partial 16S rRNA gene sequences in GenBank (DQ456336.1, DQ456227.1) [38] show 97-98% overall identity to our type e sequences by Clustal Omega. Finally, bacteria closely resembling sequence type b are, according to the MetaMetaDB-search, exclusively found in environments annotated as “human”, “gut” or “human gut”, but not in any of the 13 animal habitats, supporting our findings about the specificity of this sequence type for humans. Furthermore, the 16S rRNA gene sequence hits obtained by PrimerBlast with type b primers (with 99% identity to our type b sequences) all originate from human stool, except one that is from the feces of an orangutan.

Several recent publications have presented databases containing thousands of microbial whole genome sequences assembled from the reads of metagenomic sequencing studies [27, 39, 40]. A search against the 16S rRNA gene dataset linked to the comprehensive work of Pasolli et al [27] revealed that one of the Ct85 sequence types defined in the current work, c, has a close counterpart (99.9% 16S rRNA gene identity) within these genomic bins. The same 16S rRNA gene sequence has also previously been detected in the “metagenomic species” annotated as *Clostridium_sp_CAG_245* in the human “MetaHit” project. Two other sequence types of the current study (a and d) have counterparts with 97% 16S rRNA gene identity, and one (f) has a counterpart with 95.5% identity in the same dataset [27], indicating that metagenomics-based genome sequences tentatively representing the same genera as those represented by types a, d and f are available. Close relatives of the other Ct85 sequence types were not found in this 16S rRNA gene dataset (the closest 16S rRNA gene similarities with the sequences in the dataset were 91-92% for type b and 93-94% for type e sequences) (Supplementary Table 2).

Discussion

Sequences of the intestinal 16S rRNA gene group called the “Ct85-cluster” have been encountered in numerous previous studies, as hundreds of 16S rRNA gene sequences hybridizing with Ct85-primers[8] *in silico* and having high similarity to Ct85-cluster sequences can be found in public nucleotide databases. However, the great majority of the bacterial members of this 16S rRNA-gene group, i.e. the representatives of the sequence types a-f defined in this work, have never been cultured. While large databases containing bacterial whole genome sequences assembled from metagenomic data have recently come available [27, 39, 40], we have now detected close counterparts for the 16S rRNA genes of some of the Ct85 sequence types in these species-level genomic bins. All these genomic hits originate from human intestinal metagenomes and represent uncultivated bacteria by definition. They might, however, serve as starting points for the targeted design of isolation conditions for Ct85-cluster bacteria in the future.

Our own previous, extensive and systematic efforts to isolate Ct85-cluster representatives in pure culture, using numerous growth and enrichment media and dozens of different antibiotics as selective factors under strictly anaerobic conditions, were unsuccessful (see Supplementary Text for further details). Therefore, since the typed members of the Ct85-cluster, as defined in this study, have so far only been detected based on their 16S rRNA genes, nothing is known about their lifestyle, growth requirements or interactions within the microbial ecosystem or with the host. Before this work also the sequence heterogeneity within this cluster, as well as the prevalence and abundance of the members of this group in different human and animal populations have been completely unexplored. In this work we thus sought to provide this basic information, crucial to improve our understanding of the ecology of this large group of mostly uncultured bacteria.

438 In the current study we have shown that the Ct85-group of 16S rRNA gene sequences is much
 439 more heterogeneous than previously presumed, which is apparently due to the inadequacy of the
 440 original Ct85-targeting primers [8]: the lowest pairwise similarity between Ct85-cluster
 441 representatives, after excluding the two sequences with similarities to *Ruminococcus* or *Roseburia*,
 442 is 89.9% (82% in the original sequence panel) (Fig. 1). 82.0% has been considered as the 16S rRNA
 443 gene similarity threshold for bacteria belonging to the same order, and values 86.5%, 94.5% and
 444 98.7% as thresholds for bacteria belonging to the same family, genera or species, respectively [41].
 445 The Ct85 sequence types defined by us would thus correspond to species (sequence type b),
 446 genera (sequence types a, e and f) or families (sequence types c and d), based on both the 16S
 447 rRNA gene sequences in our sequence panel and those, of uncultured bacteria, available in public
 448 databases (see Materials and Methods). All these taxa could, based on current knowledge, be
 449 placed in one order. Currently the cultured bacteria phylogenetically closest to the Ct85 sequence
 450 types defined in this work belong to the genera *Hungateiclostridium*, *Pseudoclostridium* or
 451 *Ruminiclostridium*. Their 16S rRNA gene similarities to the newly defined Ct85 types are, however,
 452 only 86%, as determined by blastn, and even less as determined by Clustal Omega (see Figure 1A).
 453 We have previously reported differences between the abundances of Ct85-cluster 16S rRNA gene
 454 sequences in healthy individuals and IBS-M sufferers compared to those with IBS-D in the same
 455 Finnish cohort[8] as used as a part of the current study. The differences in Ct85-cluster 16S rRNA
 456 gene abundances between subject groups in the Finnish cohort were however not seen in the
 457 Dutch cohort alone, due to large interindividual variation (data not shown). In the current work we
 458 demonstrated that not all Ct85 sequence types are differentially present in the subject groups of
 459 the Finnish IBS cohort. Moreover, we demonstrated remarkable variation between the two
 460 cohorts in the Ct85 sequence type compositions across subject groups: most of the significant
 461 differences seen between groups in the Finnish cohort were not present, or not evaluable, in the

Dutch cohort. The datasets were kept separate throughout the study because the samples had not been collected using the same regimen and different diagnostic criteria for IBS (Rome II vs Rome III) had been used. The only difference in prevalence observed consistently in both cohorts was the higher occurrence of sequence type e in healthy individuals compared to subjects suffering from IBS-M; this would have been statistically significant also if the datasets had been combined, while combining the datasets did not reveal other significant differences between subject groups (data not shown). Additionally, sequence type d appeared not to be typical of IBS-D sufferers, as it was either not detected at all in this group (Finnish cohort) or it was rarely present at a hardly detectable level (Dutch cohort). However, as also revealed by the wide confidence intervals of the observed prevalences (Figure 2), the low numbers of subjects in the IBS subject groups and the low incidence of many sequence types in these groups hampered drawing conclusions from the data; more similarities between the cohorts might have been observed with a larger number of samples. The number of typed samples in the Dutch cohort was further lowered by the failure of PCR I reactions of several samples. The reason for this failure is unknown; we presume that the DNA extraction method used, though proven to faithfully maintain the original microbial diversity [21], is however rather harsh owing to its mechanical cell disruption method, and may occasionally have caused too extensive DNA fragmentation. In addition to the small sample size, a further source of uncertainty in the current work arises from the fact that despite careful primer design and testing *in silico* and *in vitro*, we cannot completely rule out the possibility of some unspecific amplification in our complex fecal DNA samples.

Our detection of more Ct85 sequence types in healthy subjects than in individuals with IBS, especially when comparing healthy controls to IBS-D sufferers, supports the generally accepted view that higher microbial diversity in the gut ecosystem provides resilience and contributes to better intestinal health[42, 43], at least in case of diarrheal diseases[44]. Fecal bacterial alpha-

486 diversity in individuals with IBS has indeed been shown to be either lower or the same, but not
487 higher than in healthy controls[45]. Interestingly, in the study of Turnbaugh et al[35], in which the
488 gut microbiome of lean and obese twin pairs was investigated by metagenomic methods,
489 numerous Ct85-cluster sequences representing sequence types e and f were detected, and they
490 were exclusively found in the fecal samples of lean rather than obese or overweight individuals.
491 Similarly, Ferrer et al found a c-type Ct85 sequence (JX543478.1) only in the lean subject when
492 investigating the microbiotas of a single lean and obese individual by metagenomic methods [46].
493 These data support the view of at least some sequence types of Ct85 as bacteria indicative of
494 general intestinal and metabolic health.

495 Before this work, the presence of representatives of the Ct85-cluster had not been systematically
496 studied in animals. Our small sample material did not allow for extensive conclusions, but some
497 trends could be recognized. In addition to humans, Ct85-cluster members seemed to be common
498 in pigs, which are omnivores having a similar gastrointestinal system as humans, and are therefore
499 frequently used to model human digestive tract function[47]. Even the Ct85 sequence type profile
500 of pigs resembled that of humans in our study, as the same types (excluding type b not found in
501 any animal) were present in both species at approximately same relative abundances, and the
502 most common sequence type, f, was the same in both humans and adult swine. In our sample
503 material of humans and animals, sequence type b appeared as human specific. A survey of public
504 databases revealed the presence of this sequence type additionally in apes, suggesting specificity
505 of this putative species-level sequence type for primates. Sequence type d, in turn, was in our
506 study rather uncommon in humans while very frequently found in swine and reared wild boars in
507 high numbers, but not in any other animals we tested. These findings were supported by public
508 database searches, which revealed sequences closely resembling type d in pig, human and ape
509 stool, but also commonly in other environmental, mostly anaerobic habitats, suggesting that the

human gut is not the primary habitat for this sequence type. Furthermore, sequence type d was present in each suckling 1 week old piglet and in each sow-piglet pair tested, suggesting maternal transmission of this type in pigs. The same was true for sequence type f, as suckling 1 week old piglets mostly carried this type of Ct85 in their guts like their mothers, while the other sequence types detected in swine, a, c and e, did not so clearly seem to be maternally transmitted. We detected the two common sequence types, e and f, in high numbers also in single fecal samples of two laboratory animals, rat and mouse, in which they represented the only Ct85 sequence types detectable (data not shown).

In this work, we presented a method to type the 16S rRNA genes of the Ct85-cluster by PCR. Then we mapped the distribution of these Ct85 sequence types in human and animal populations, and the distribution of the Ct85-cluster sequences in animals, and showed that while Ct85-cluster sequences are present in several species, the composition and abundance of Ct85 sequence types differ remarkably between animal species. These kind of data are crucial to improve our understanding of the ecology of uncultured bacteria like Ct85-cluster members, and they also serve as a basis for further, directed culturing efforts based on e.g. single cell isolation and genomic sequencing, followed by tailoring isolation media based on genomic data now available for some sequence types (see above) [48-50]. The Ct85-cluster is an especially interesting group of uncultured bacteria due to its ubiquity and anticipated role in human intestinal health. Further studies with larger numbers of samples, comparing the presence of Ct85 sequence types across human or animal populations or between IBS sufferers and healthy individuals, are however warranted to facilitate future efforts to isolate representatives of especially those Ct85 sequence types that are diagnostic for healthiness or for a particular IBS-type. Culturing and genomic sequencing, followed by knowledge of the ecology and interactions of Ct85 sequence types, may

then promote our understanding of the microbial component of IBS etiology and help in subject stratification, diagnosis and cure.

Conclusions

In this study, the large and almost completely uncultured intestinal bacterial group called the Ct85-cluster, previously identified as non-IBS-D –associated and supposed to form a coherent phylogenetic group, was now demonstrated to comprise a heterogeneous group of bacteria. Six sequence types of Ct85 distinguishable by PCR were identified, each representing a novel, uncultivated genus-, species- or family-level taxon based on 16S rRNA gene similarities. Ct85-cluster members were found to be present in multiple companion and production animal species, but the sequence type prevalences and abundances were drastically different across species. The sequence type composition of adult swine closely resembled that of humans. One of the newly identified types, tentatively corresponding to a species, was in this study only detected in humans (b), while one of the family-level types shared between humans and swine (d) was not detected in any other animal tested. By using two cohorts of IBS sufferers and healthy controls from two countries, we demonstrated that one of the tentative genus-level sequence types (e) was more prevalent in healthy subjects than in those with IBS-M, but no other differences observed in one cohort could be reiterated in the other, emphasizing geographic and individual variation in IBS etiology and the need for larger study cohorts. Typing of microbial groups, such as the Ct85-cluster, so far almost only detectable by their 16S rRNA genes, is a necessary initial step to improve our understanding of the ecology of uncultured taxons. Future efforts, aiming at pure culturing of Ct85-cluster representatives and revealing their interactions with their hosts are strongly warranted, and studies focused on Ct85 sequence types differentially present in healthy subjects and different IBS types will be of special interest.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by the Jane and Aatos Erkko Foundation and the Academy of Finland (grant number 277362). Anja Osola is thanked for skilled technical assistance, Dr Gerben Hermes for help with the samples of the Dutch IBS cohort, which was supported by the unrestricted Spinoza 2018 award of the Netherlands Organization for Scientific Research (NWO to WMdV). Dr Per Johansson is thanked for advice with bioinformatics, DVM Ida Piispa for organizing the collection of dog and cat fecal samples, the personnel in the Viikki Laboratory Animal Center for providing the rodent fecal samples and Dr Miika Kahelin for the opportunity to collect bovine fecal samples.

Table 1. Numbers of the human fecal samples analyzed

Cohort	Subject group	Number of fecal samples		
		Initial	Ct85 positive	Typed
Finnish [6, 7]	Healthy	23	23	23
	IBS-D	12	12	12
	IBS-M	6	6	6
	IBS-C	9	9	9
	Total	50	50	50
Dutch (manuscript in preparation)	Healthy	51	48	44
	IBS-D	13	13	5
	IBS-M	14	14	8
	IBS-C	21	21	11
	IBS-type not known	1	1	1
	Total	100	97	69

Table 2. Primer sequences specific for the Ct85 sequence types and optimized conditions for end-point PCR II.

Ct85 sequence type	Primer sequences	Annealing temperature (°C)	Extension time (s)	Amplicon size (bp)
a	F: 5' GCTTTAGACAGGGAAGAACAAG 3' R: 5' GGGATTTCACTAACTTGCATTA 3'	68	3	155
b	F: 5' GTGATTAGTTACTAACAAGTTAAGTTGAGG 3' R: 5' TGCATCACTGCTTGCCT 3'	67	3	158
c	F: 5' GATGAATAACTAAGAGATTAGTTAGTC 3' R: 5' TAACCTGTTGGCACTAATC 3'	58	2	137
d	F: 5' AGGTGTAGGGAACGATATGTTT 3' R: 5' ACAAGTCTCGAAAGACTACCTAATT 3'	62	4	214
e	F: 5' GCAACACAGTGATGTGAAG 3' R: 5' CTCTGAAAAGTTGGCTCAC 3'	65	3	200
f	F: 5' ATATTTCTTGAGTRCAGGAG 3' R: 5' CTATATAATCTCTTATATATTCATCTAT 3'	61	6	380

Table 3. Prevalences and abundances of Ct85-cluster 16S rRNA genes in animal fecal samples. The prevalences are expressed as the percentage of animals with Ct85-cluster 16S rRNA genes detected in the fecal sample. CI, confidence interval.

Animal (n)	Prevalence % (95% CI)	Abundance of Ct85-cluster 16S rRNA genes					
		Copies / g feces			Proportion of total 16S rRNA gene copies		
		Range	Median	Median of positive values	Range	Median	Median of positive values
Sow (25)	100 (87-100)	1.0×10^8 - 9.2×10^9	1.3×10^9	1.3×10^9	4.8×10^{-4} - 2.3×10^{-2}	3.7×10^{-3}	3.7×10^{-3}
Piglet (7)	100 (65-100)	8.7×10^6 - 2.1×10^9	9.1×10^8	9.1×10^8	7.7×10^{-6} - 4.3×10^{-4}	1.4×10^{-4}	1.4×10^{-4}
Wild boar (9)	100 (70-100)	4.6×10^6 - 5.5×10^7	1.8×10^7	1.8×10^7	3.8×10^{-5} - 4.1×10^{-4}	6.2×10^{-5}	6.2×10^{-5}
Dog (19)	63 (41-81)	1.3×10^3 - 5.1×10^7	4.3×10^3	6.3×10^4	2.2×10^{-8} - 1.0×10^{-3}	2.1×10^{-7}	8.6×10^{-7}
Cat (17)	82 (59-94)	7.5×10^2 - 7.5×10^7	3.7×10^4	7.0×10^4	2.3×10^{-9} - 3.2×10^{-4}	6×10^{-8}	7.5×10^{-8}
Cow (12)	100 (76-100)	8.8×10^7 - 1.4×10^9	7.9×10^8	7.9×10^8	1.5×10^{-4} - 2.5×10^{-3}	1.3×10^{-3}	1.3×10^{-3}

576

577 **Table 4.** Prevalences of Ct85 sequence types in animal fecal samples determined by end-point

578 PCR, expressed as the percentage of animals with the particular sequence type in the fecal sample.

579 †, samples end-point PCR -positive, but the copy numbers in some weakly positive samples (one

580 sow, three piglet samples) remained below the qPCR detection limit. CI, confidence interval.

Animal (n)	Prevalence of Ct85 sequence type, % (95% CI)					
	a	b	c	D	e	f
Sow (14)	86 (60-96)	0 (0-22)	36 (16-61)	71 (45-88)	86 (60-96) [†]	100 (78-100)
Piglet (6)	33 (10-70)	0 (0-39)	17 (3-56) [†]	100 (61-100)	67 (30-90) [†]	67 (30-90)
Wild boar (2)	100 (34-100)	0 (0-66)	0 (0-66)	100 (34-100)	100 (34-100)	50 (9-91)
Dog (11)	9 (2-38)	0 (0-26)	0 (0-26)	0 (0-26)	18 (5-48)	9 (2-38)
Cat (12)	8 (1-35)	0 (0-24)	0 (0-24)	0 (0-24)	67 (39-86)	25 (9-53)

581

582

Figure Legends

Figure 1. (A) Pairwise similarities of the near full length Ct85-cluster 16S rRNA gene sequences, originating from [7] and used in this study for defining Ct85 sequence types, and of their closest cultured relatives, determined by Clustal Omega. The sequence codes refer to the clone names under ENA accession numbers LR595726-595746; for abbreviations of the cultured species, see Fig. 1B. The sequences belonging to each type and the corresponding similarity values are framed. The Ct85 sequence type designations (a-f) are shown on the left. (B) A maximum likelihood phylogenetic tree based on the cloned, near full-length 16S rRNA gene sequences of the Ct85-cluster, compared in panel A, and the 16S rRNA gene sequences of their closest cultured relatives. The tree was constructed by MEGA7; of the initial trees created automatically, the one with the highest log likelihood is shown. Scale bar indicates the number of nucleotide substitutions per site.

Figure 2. Prevalences of Ct85 sequence types a to f in healthy individuals and in different IBS symptom groups of the Finnish (A) and Dutch (B) study cohort, expressed as the percentage of individuals in the group with the particular sequence type in the fecal sample. Whiskers represent 95% confidence intervals. Sample numbers in each group (n) are in parentheses. Statistically significant differences between groups are indicated by lines; *, $P \leq 0.05$. The false discovery rate-corrected P-values for the Fisher's exact tests assessing the association between the Ct85 sequence type and the subject groups are as follows: Finnish cohort: 0.919, 0.052, 0.039, 0.039, 0.024 and 0.024 for sequence types a, b, c, d, e and f, respectively; Dutch cohort: 0.840, 0.840, 0.840, 0.840, 0.618 and 0.840 for sequence types a, b, c, d, e and f, respectively.

Figure 3. Relative abundances of the Ct85 sequence types a to e, determined by qPCR with type-specific and universal primers and expressed as proportions of the 16S rRNA genes representing different sequence types of total, in the fecal samples of the subjects of the Finnish (black circles) and Dutch (white circles) IBS study cohorts, and in the fecal samples of animals (blue-rimmed

circles). Each circle represents the relative abundance of the sequence type in the fecal sample of a single human or animal. Letters in the panels correspond to the naming of the sequence types (a to e; setting up a qPCR assay for sequence type f failed). The prevalences of the sequence types in the animal groups are shown in parentheses.

Supplementary Information

Supplementary Table 1 (docx). Presence of Ct85 sequence types (a to f) in the fecal samples of six sow/piglet pairs determined by end-point PCR.

Supplementary Table 2 (xlsx). Top blastn hits for the cloned Ct85-cluster 16S rRNA gene sequences in the SGB dataset of Pasolli et al [27].

Supplementary Figure 1 (pptx). Numbers of Ct85 sequence types in the fecal samples of healthy individuals and IBS sufferers in the Finnish (left, A) and Dutch (right, B) study cohorts. In panels A1 and B1, medians, 25 to 75 percentiles, minimum and maximum values and an outlier are shown as horizontal black lines, boxes, whiskers and a point, respectively. In panels A2-A6 and B2-B6, the numbers of sequence types are expressed as the numbers (bars) or proportions (lines) of individuals with a particular number of Ct85 sequence types in the fecal sample. A2 and B2, healthy subjects; A3 and B3, IBS sufferers; A4 and B4, IBS-D sufferers; A5 and B5 IBS-M sufferers, and A6 and B6, IBS-C sufferers.

Supplementary Figure 2 (pptx). Results of the MetaMetaDB search [25] performed for the 16S rRNA gene sequences representing Ct85 sequence types a-f. The sequence type designations are shown in parentheses. The y-axis shows the Microbial Habitability Indices (MHI) of the environments listed below the graphs and marked by different colors. The columns show the summary of the MHIs calculated by those BLAST hits that are above the identity threshold shown

629 below the column. The environments, the MHIs of which are less than 1% in each column, are
630 summed up and labeled as "other."

631 **Supplementary Text (docx).** A description of the cultivation efforts of Ct85-cluster
632 representatives.

Journal Pre-proof

References

- 633
- 634 1. Canavan C, West J, Card T: **The epidemiology of irritable bowel syndrome.** *Clin Epidemiol* 2014, **6**:71-80.
- 635 2. Enck P, Aziz Q, Barbara G, Farmer AD, Fukudo S, Mayer EA, Niesler B, Quigley EM, Rajilic-Stojanovic M,
- 636 Schemann M, Schwille-Kiuntke J, Simren M, Zipfel S, Spiller RC: **Irritable bowel syndrome.** *Nat Rev Dis*
- 637 *Primers* 2016, **2**:16014.
- 638 3. Öhman L, Tornblom H, Simren M: **Crosstalk at the mucosal border: importance of the gut**
- 639 **microenvironment in IBS.** *Nat Rev Gastroenterol Hepatol* 2015, **12**(1):36-49.
- 640 4. Duan R, Zhu S, Wang B, Duan L: **Alterations of Gut Microbiota in Patients With Irritable Bowel**
- 641 **Syndrome Based on 16S rRNA-Targeted Sequencing: A Systematic Review.** *Clin Transl Gastroenterol* 2019,
- 642 **10**(2):e00012.
- 643 5. Carabotti M, Scirocco A, Maselli MA, Severi C: **The gut-brain axis: interactions between enteric**
- 644 **microbiota, central and enteric nervous systems.** *Ann Gastroenterol* 2015, **28**(2):203-209.
- 645 6. Malinen E, Rinttilä T, Kajander K, Mättö J, Kassinen A, Krogus L, Saarela M, Korpela R, Palva A: **Analysis of**
- 646 **the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR.** *Am J*
- 647 *Gastroenterol* 2005, **100**(2):373-382.
- 648 7. Kassinen A, Krogus-Kurikka L, Mäkiuokko H, Rinttilä T, Paulin L, Corander J, Malinen E, Apajalahti J,
- 649 Palva A: **The fecal microbiota of irritable bowel syndrome patients differs significantly from that of**
- 650 **healthy subjects.** *Gastroenterology* 2007, **133**(1):24-33.
- 651 8. Lyra A, Rinttilä T, Nikkilä J, Krogus-Kurikka L, Kajander K, Malinen E, Mättö J, Mäkelä L, Palva A:
- 652 **Diarrhoea-predominant irritable bowel syndrome distinguishable by 16S rRNA gene phylotype**
- 653 **quantification.** *World J Gastroenterol* 2009, **15**(47):5936-5945.
- 654 9. Hynönen U, Rasinkangas P, Satokari R, Paulin L, de Vos WM, Pietilä TE, Kant R, Palva A: **Isolation and**
- 655 **whole genome sequencing of a *Ruminococcus*-like bacterium, associated with irritable bowel syndrome.**
- 656 *Anaerobe* 2016, **39**:60-67.
- 657 10. Seo B, Yoo JE, Lee YM, Ko G: ***Sellimonas intestinalis* gen. nov., sp. nov., isolated from human faeces.** *Int*
- 658 *J Syst Evol Microbiol* 2016, **66**(2):951-956.
- 659 11. Lyra A, Krogus-Kurikka L, Nikkilä J, Malinen E, Kajander K, Kurikka K, Korpela R, Palva A: **Effect of a**
- 660 **multispecies probiotic supplement on quantity of irritable bowel syndrome-related intestinal microbial**
- 661 **phylotypes.** *BMC Gastroenterol* 2010, **10**:110-230X-10-110.
- 662 12. Kajander K, Hatakka K, Poussa T, Färkkilä M, Korpela R: **A probiotic mixture alleviates symptoms in**
- 663 **irritable bowel syndrome patients: a controlled 6-month intervention.** *Aliment Pharmacol Ther* 2005,
- 664 **22**(5):387-394.
- 665 13. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for**
- 666 **Bigger Datasets.** *Mol Biol Evol* 2016, **33**(7):1870-1874.
- 667 14. Tamura K, Nei M: **Estimation of the number of nucleotide substitutions in the control region of**
- 668 **mitochondrial DNA in humans and chimpanzees.** *Mol Biol Evol* 1993, **10**(3):512-526.

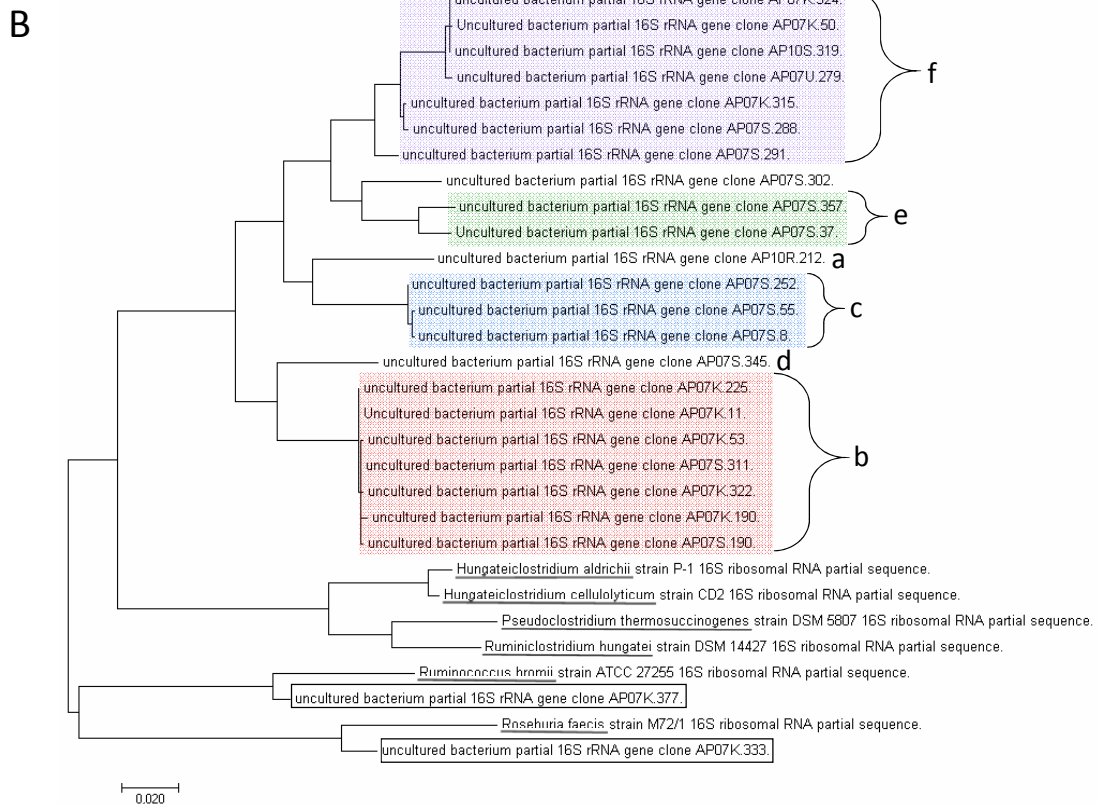
- 669 15. Maukonen J, Satokari R, Mättö J, Söderlund H, Mattila-Sandholm T, Saarela M: **Prevalence and**
 670 **temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant**
 671 **faecal bacteria.** *J Med Microbiol* 2006, **55**(Pt 5):625-633.
- 672 16. Malinen E, Rinttilä T, Kajander K, Mättö J, Kassinen A, Krogus L, Saarela M, Korpela R, Palva A: **Analysis**
 673 **of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR.** *Am*
 674 *J Gastroenterol* 2005, **100**(2):373-382.
- 675 17. Krogus-Kurikka L, Lyra A, Malinen E, Aarnikunnas J, Tuimala J, Paulin L, Mäkiyuokko H, Kajander K, Palva
 676 A: **Microbial community analysis reveals high level phylogenetic alterations in the overall gastrointestinal**
 677 **microbiota of diarrhoea-predominant irritable bowel syndrome sufferers.** *BMC Gastroenterol* 2009, **9**:95-
 678 230X-9-95.
- 679 18. Kajander K, Krogus-Kurikka L, Rinttilä T, Karjalainen H, Palva A, Korpela R: **Effects of multispecies**
 680 **probiotic supplementation on intestinal microbiota in irritable bowel syndrome.** *Aliment Pharmacol Ther*
 681 2007, **26**(3):463-473.
- 682 19. Mättö J, Maunuksela L, Kajander K, Palva A, Korpela R, Kassinen A, Saarela M: **Composition and**
 683 **temporal stability of gastrointestinal microbiota in irritable bowel syndrome--a longitudinal study in IBS**
 684 **and control subjects.** *FEMS Immunol Med Microbiol* 2005, **43**(2):213-222.
- 685 20. Hasan S, Junnikkala S, Peltoniemi O, Paulin L, Lyyski A, Vuorenmaa J, Oliviero C: **Dietary**
 686 **supplementation with yeast hydrolysate in pregnancy influences colostrum yield and gut microbiota of**
 687 **sows and piglets after birth.** *PLoS One* 2018, **13**(5):e0197586.
- 688 21. Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Nikkilä J, Palva A, de Vos WM: **Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery**
 689 **of bacterial and archaeal DNA using mechanical cell lysis.** *J Microbiol Methods* 2010, **81**(2):127-134.
- 690 22. Jakava-Viljanen M, Palva A: **Isolation of surface (S) layer protein carrying *Lactobacillus* species from**
 691 **porcine intestine and faeces and characterization of their adhesion properties to different host tissues.**
 692 *Vet Microbiol* 2007, **124**(3-4):264-273.
- 693 23. Nadkarni MA, Martin FE, Jacques NA, Hunter N: **Determination of bacterial load by real-time PCR using**
 694 **a broad-range (universal) probe and primers set.** *Microbiology* 2002, **148**(Pt 1):257-266.
- 695 24. Dore J, Sghir A, Hannequart-Gramet G, Corthier G, Pochart P: **Design and evaluation of a 16S rRNA-**
 696 **targeted oligonucleotide probe for specific detection and quantitation of human faecal *Bacteroides***
 697 **populations.** *Syst Appl Microbiol* 1998, **21**(1):65-71.
- 698 25. Sriswasdi S, Yang CC, Iwasaki W: **Generalist species drive microbial dispersion and evolution.** *Nat*
 699 *Commun* 2017, **8**(1):1162-017-01265-1.
- 700 26. Lagkouravdos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, Clavel T: **IMNGS: A**
 701 **comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies.**
 702 *Sci Rep* 2016, **6**:33721.
- 703 27. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P,
 704 Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N: **Extensive**
 705 **Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes**
 706 **Spanning Age, Geography, and Lifestyle.** *Cell* 2019, **176**(3):649-662.e20.

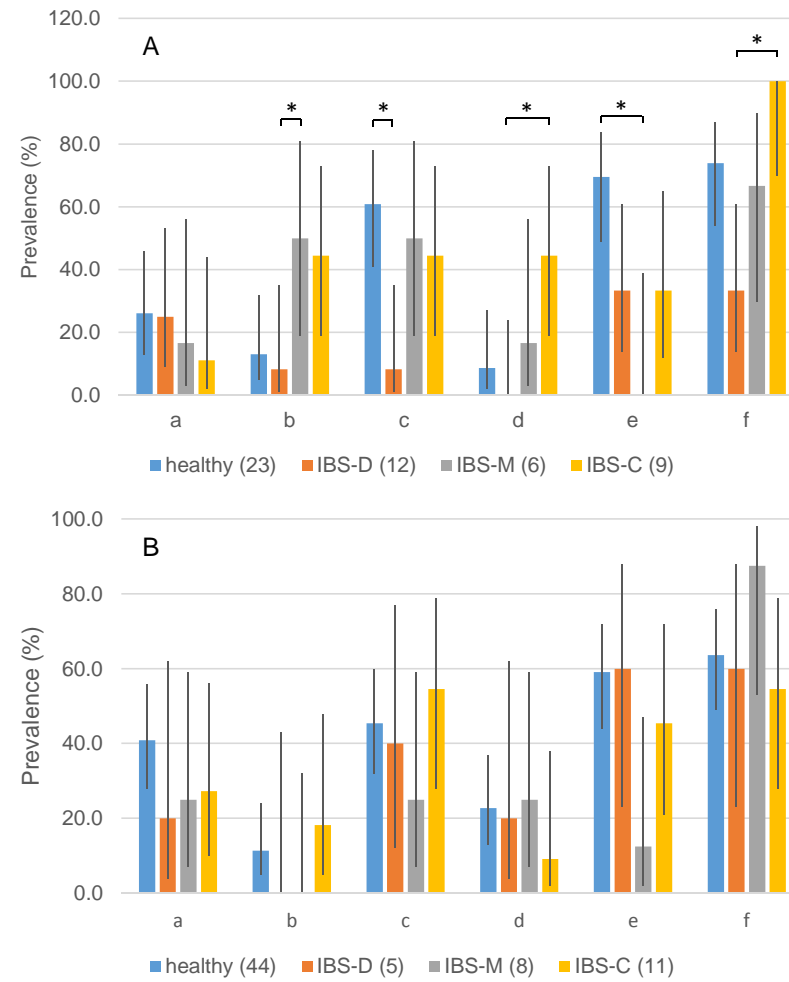
- 708 28. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA: **Database indexing for**
709 **production MegaBLAST searches.** *Bioinformatics* 2008, **24**(16):1757-1764.
- 710 29. Brown L, Cai T, DasGupta A, Agresti A, Coull B, Casella G, Corcoran C, Mehta C, Ghosh M, Santner T,
711 Brown L, Cai T, DasGupta A: **Interval estimation for a binomial proportion - Comment - Rejoinder.** *Stat Sci*
712 2001, **16**(2):101-133.
- 713 30. Mai V, Greenwald B, Morris JG, Jr, Raufman JP, Stine OC: **Effect of bowel preparation and colonoscopy**
714 **on post-procedure intestinal microbiota composition.** *Gut* 2006, **55**(12):1822-1823.
- 715 31. Lin A, Bik EM, Costello EK, Dethlefsen L, Haque R, Relman DA, Singh U: **Distinct distal gut microbiome**
716 **diversity and composition in healthy children from Bangladesh and the United States.** *PLoS One* 2013,
717 **8**(1):e53838.
- 718 32. Ley RE, Turnbaugh PJ, Klein S, Gordon JI: **Microbial ecology: human gut microbes associated with**
719 **obesity.** *Nature* 2006, **444**(7122):1022-1023.
- 720 33. Täubel M, Rintala H, Pitkäranta M, Paulin L, Laitinen S, Pekkanen J, Hyvärinen A, Nevalainen A: **The**
721 **occupant as a source of house dust bacteria.** *J Allergy Clin Immunol* 2009, **124**(4):834-40.e47.
- 722 34. Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspith BN, Rayment N, Brostoff J, Parkhill J,
723 Dougan G, Petrovska L: **High-throughput clone library analysis of the mucosa-associated microbiota**
724 **reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in**
725 **inflammatory bowel disease.** *BMC Microbiol* 2011, **11**:7-2180-11-7.
- 726 35. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA,
727 Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI: **A core gut microbiome in obese and**
728 **lean twins.** *Nature* 2009, **457**(7228):480-484.
- 729 36. Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Zhang M, Wei H, Chen
730 Y, Lu H, Zuo J, Su M, Qiu Y, Jia W, Xiao C, Smith LM, Yang S, Holmes E, Tang H, Zhao G, Nicholson JK, Li L,
731 Zhao L: **Symbiotic gut microbes modulate human metabolic phenotypes.** *Proc Natl Acad Sci U S A* 2008,
732 **105**(6):2117-2122.
- 733 37. Kalmokoff M, Waddington LM, Thomas M, Liang KL, Ma C, Topp E, Dandurand UD, Letellier A, Matias F,
734 Brooks SP: **Continuous feeding of antimicrobial growth promoters to commercial swine during the**
735 **growing/finishing phase does not modify faecal community erythromycin resistance or community**
736 **structure.** *J Appl Microbiol* 2011, **110**(6):1414-1425.
- 737 38. Scupham AJ: **Succession in the intestinal microbiota of preadolescent turkeys.** *FEMS Microbiol Ecol*
738 2007, **60**(1):136-147.
- 739 39. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD: **A new**
740 **genomic blueprint of the human gut microbiota.** *Nature* 2019, .
- 741 40. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC: **New insights from uncultivated genomes of the**
742 **global human gut microbiome.** *Nature* 2019, **568**(7753):505-510.
- 743 41. Rossello-Mora R, Amann R: **Past and future species definitions for *Bacteria* and *Archaea*.** *Syst Appl*
744 *Microbiol* 2015, **38**(4):209-216.

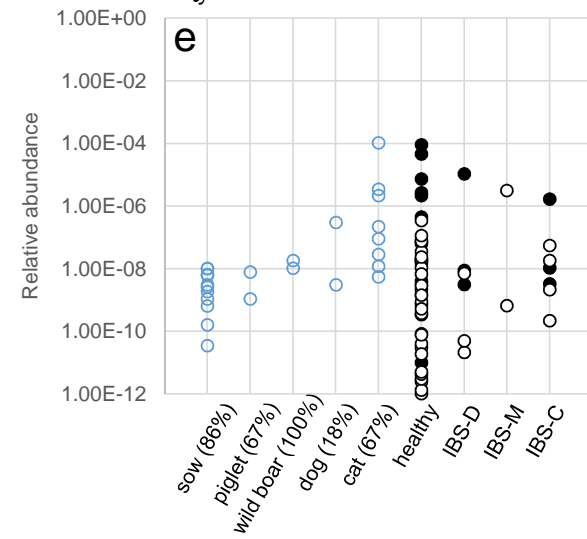
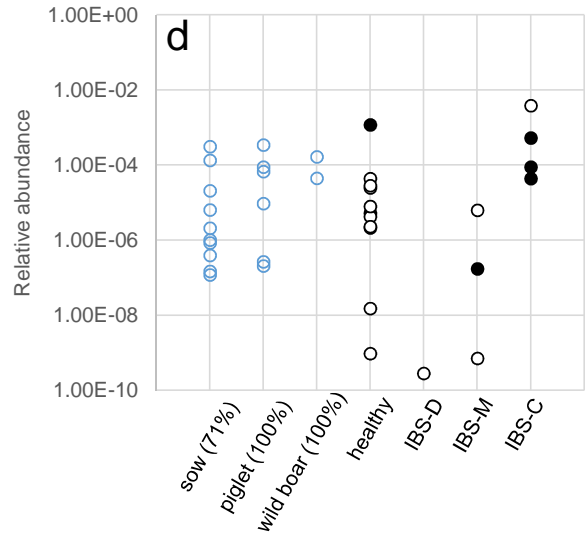
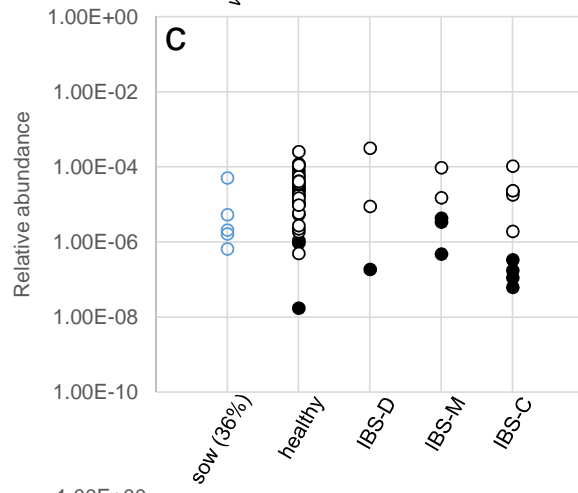
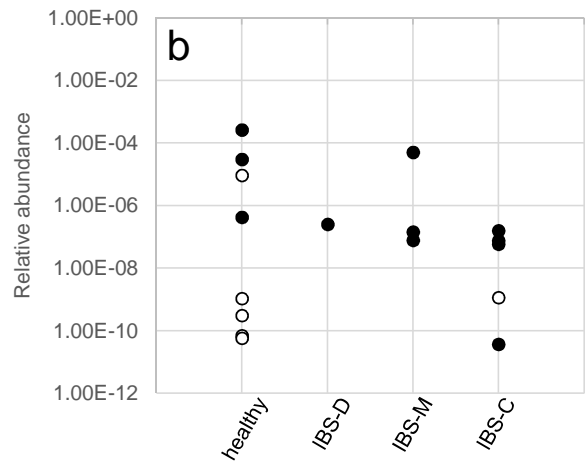
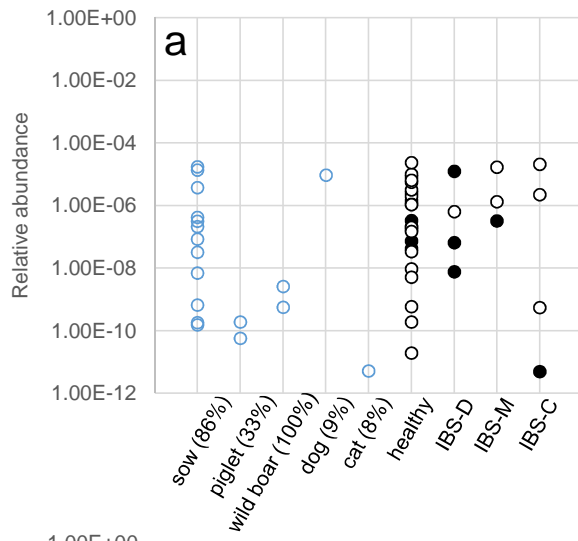
- 745 42. Bennet SM, Öhman L, Simren M: **Gut microbiota as potential orchestrators of irritable bowel**
 746 **syndrome.** *Gut Liver* 2015, **9**(3):318-331.
- 747 43. McCann KS: **The diversity-stability debate.** *Nature* 2000, **405**(6783):228-233.
- 748 44. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ: **Meta-analysis of gut microbiome studies identifies**
 749 **disease-specific and shared responses.** *Nat Commun* 2017, **8**(1):1784-017-01973-8.
- 750 45. Pittayanon R, Lau JT, Yuan Y, Leontiadis GI, Tse F, Surette M, Moayyedi P: **Gut Microbiota in Patients**
 751 **With Irritable Bowel Syndrome-a Systematic Review.** *Gastroenterology* 2019, in press.
- 752 46. Ferrer M, Ruiz A, Lanza F, Haange SB, Oberbach A, Till H, Bargiela R, Campoy C, Segura MT, Richter M,
 753 von Bergen M, Seifert J, Suarez A: **Microbiota from the distal guts of lean and obese adolescents exhibit**
 754 **partial functional redundancy besides clear differences in community structure.** *Environ Microbiol* 2013,
 755 **15**(1):211-226.
- 756 47. Hillman ET, Lu H, Yao T, Nakatsu CH: **Microbial Ecology along the Gastrointestinal Tract.** *Microbes*
 757 *Environ* 2017, **32**(4):300-313.
- 758 48. Gutleben J, Chaib De Mares M, van Elsas JD, Smidt H, Overmann J, Sipkema D: **The multi-omics promise**
 759 **in context: from sequence to microbial isolate.** *Crit Rev Microbiol* 2018, **44**(2):212-229.
- 760 49. Overmann J, Abt B, Sikorski J: **Present and Future of Culturing Bacteria.** *Annu Rev Microbiol* 2017,
 761 **71**:711-730.
- 762 50. Allen-Vercoe E: **Bringing the gut microbiota into focus through microbial culture: recent progress and**
 763 **future perspective.** *Curr Opin Microbiol* 2013, **16**(5):625-629.

764

A	AP07K.333	<i>R. faecis</i>	AP07K.377	<i>R. bromii</i>	AP10R.212	AP07S.252	AP07S.55	AP07S.8	AP07S.357	AP07S.37	AP07S.302	AP07S.291	AP07K.315	AP07S.288	AP07U.279	AP10S.319	AP07K.50	AP07K.324	AP07S.345	AP07K.190	AP07K.53	AP07S.190	AP07K.322	AP07S.311	AP07K.11	AP07K.225	<i>P. thermos.</i>	<i>R. hungatei</i>	<i>H. aldrichii</i>	<i>H. cellulol.</i>		
	AP07K.333	100.0	92.31	81.97	78.69	83.59	84.34	84.20	84.10	84.40	83.82	82.20	82.80	84.83	85.12	86.26	86.39	86.27	86.53	83.53	82.21	81.86	82.21	82.21	82.21	81.95	82.21	79.07	79.37	81.42	81.67	
	<i>R. faecis</i>	92.31	100.0	78.67	78.77	79.05	80.25	80.11	80.25	78.93	78.41	78.92	78.71	79.00	78.86	79.06	79.13	79.06	79.13	79.73	78.84	78.77	78.70	78.77	78.77	78.77	78.77	79.34	79.00	80.73	81.06	
	AP07K.377	81.97	78.67	100.0	92.31	81.49	82.94	82.74	82.60	82.94	82.28	81.80	82.08	82.65	82.85	82.65	82.92	82.67	82.92	83.65	85.93	86.00	85.93	86.14	86.07	85.87	86.00	79.75	79.56	82.02	82.14	
	<i>R. bromii</i>	78.69	78.77	92.31	100.0	78.51	78.92	78.70	78.77	78.18	78.52	78.81	78.66	79.09	79.02	78.95	79.09	79.02	79.09	77.53	78.35	78.64	78.43	78.57	78.50	78.43	78.43	80.42	79.77	81.39	81.87	
a	AP10R.212	83.59	79.05	81.49	78.51	100.0	91.97	91.84	91.88	91.38	91.38	91.72	93.30	92.28	92.42	92.36	92.16	91.98	92.16	89.93	89.80	89.87	89.99	89.93	90.06	89.92	90.06	81.99	82.46	83.28	83.31	
	AP07S.252	84.34	80.25	82.94	78.92	91.97	100.0	99.80	99.87	93.43	92.46	92.36	91.72	92.09	91.97	91.71	91.97	91.79	91.97	91.83	91.36	91.40	91.56	91.50	91.63	91.58	91.69	81.46	82.34	84.05	84.03	
c	AP07S.55	84.20	80.11	82.74	78.70	91.84	99.80	100.0	99.87	93.30	92.33	92.23	91.59	92.02	91.84	91.58	91.84	91.66	91.84	91.69	91.23	91.26	91.43	91.36	91.50	91.44	91.56	81.40	82.14	83.98	83.96	
	AP07S.8	84.10	80.25	82.60	78.77	91.88	99.80	99.87	100.0	93.24	92.35	92.30	91.49	92.02	91.75	91.62	91.88	91.82	91.88	91.73	91.13	91.26	91.33	91.26	91.40	91.47	91.47	81.38	81.99	83.89	84.04	
e	AP07S.357	84.40	78.93	82.94	78.18	91.38	93.43	93.30	93.24	100.0	97.20	93.31	93.50	94.17	94.28	94.15	94.28	94.13	94.28	92.29	90.77	90.73	90.97	90.90	91.03	90.91	91.03	81.00	81.27	83.92	84.10	
	AP07S.37	83.82	78.41	82.28	78.52	91.38	92.46	92.33	92.35	97.20	100.0	93.77	93.47	94.30	94.33	94.06	94.19	94.13	94.19	91.16	90.29	90.38	90.50	90.43	90.56	90.56	90.56	81.74	81.53	84.05	84.55	
	AP07S.302	82.20	78.92	81.80	78.81	91.72	92.36	92.23	92.30	93.31	93.77	100.0	95.57	93.30	93.17	92.97	93.03	92.97	93.03	89.87	90.34	90.47	90.54	90.47	90.60	90.54	90.54	80.68	80.46	81.68	82.10	
	AP07S.291	82.80	78.71	82.08	78.66	93.30	91.72	91.59	91.49	93.50	93.47	95.57	100.0	96.92	97.11	95.47	95.47	95.34	95.47	90.32	90.45	90.40	90.65	90.59	90.72	90.59	90.72	81.09	80.96	82.15	82.37	
	AP07K.315	84.83	79.00	82.65	79.09	92.28	92.09	92.02	92.02	94.17	94.30	93.30	96.92	100.0	99.73	97.59	97.66	97.59	97.66	90.71	90.38	90.51	90.58	90.51	90.65	90.65	90.65	81.63	81.56	83.03	83.46	
f	AP07S.288	85.12	78.86	82.85	79.02	92.42	91.97	91.84	91.75	94.28	94.33	93.17	97.11	99.73	100.0	97.57	97.63	97.53	97.63	90.89	90.56	90.51	90.76	90.69	90.83	90.70	90.83	81.70	81.78	83.21	83.45	
	AP07U.279	86.26	79.06	82.65	78.95	92.36	91.71	91.58	91.62	94.15	94.06	92.97	95.47	97.59	97.57	100.0	99.54	99.47	99.54	90.56	90.17	90.11	90.23	90.30	90.30	90.16	90.30	81.77	82.52	84.10	84.21	
	AP10S.319	86.39	79.13	82.92	79.09	92.16	91.97	91.84	91.88	94.28	94.19	93.03	95.47	97.66	97.63	99.54	100.0	99.80	99.87	90.83	90.30	90.24	90.36	90.43	90.43	90.29	90.43	81.77	82.52	84.03	84.28	
	AP07K.50	86.27	79.06	82.67	79.02	91.98	91.79	91.66	91.82	94.13	94.13	92.97	95.34	97.59	97.53	99.47	99.80	100.0	99.93	90.63	90.09	90.17	90.16	90.23	90.23	90.23	90.23	81.73	82.34	83.82	84.25	
	AP07K.324	86.53	79.13	82.92	79.09	92.16	91.97	91.84	91.88	94.28	94.19	93.03	95.47	97.66	97.63	99.54	99.87	99.93	100.0	90.83	90.30	90.24	90.36	90.43	90.43	90.29	90.43	81.90	82.59	84.10	84.35	
d	AP07S.345	83.53	79.73	83.65	77.53	89.93	91.83	91.69	91.73	92.29	91.16	89.87	90.32	90.71	90.89	90.56	90.83	90.63	90.83	100.0	93.23	93.16	93.16	93.29	93.29	93.27	93.36	81.05	81.75	83.45	83.34	
	AP07K.190	82.21	78.84	85.93	78.35	89.80	91.36	91.23	91.13	90.77	90.29	90.34	90.45	90.38	90.56	90.17	90.30	90.09	90.30	93.23	100.0	99.73	99.61	99.74	99.74	99.67	99.67	81.53	82.36	83.12	83.29	
	AP07K.53	81.86	78.77	86.00	78.64	89.87	91.40	91.26	91.26	90.73	90.38	90.47	90.40	90.51	90.51	90.11	90.24	90.17	90.24	93.16	99.73	100.0	99.73	99.87	99.87	99.80	99.80	81.52	82.29	83.01	83.37	
b	AP07S.190	82.21	78.70	85.93	78.43	89.99	91.56	91.43	91.33	90.97	90.50	90.54	90.65	90.58	90.76	90.23	90.36	90.16	90.36	93.16	99.61	99.73	100.0	99.74	99.87	99.87	99.80	99.80	81.73	82.57	83.05	83.22
	AP07K.322	82.21	78.77	86.14	78.57	89.93	91.50	91.36	91.26	90.90	90.43	90.47	90.59	90.51	90.69	90.30	90.43	90.23	90.43	93.29	99.74	99.87	99.74	100.0	99.87	99.80	99.80	81.66	82.50	83.25	83.43	
	AP07S.311	82.21	78.77	86.07	78.50	90.06	91.63	91.50	91.40	91.03	90.56	90.60	90.72	90.65	90.83	90.30	90.43	90.23	90.43	93.29	99.74	99.87	99.87	99.87	100.0	99.93	99.93	81.73	82.57	83.19	83.36	
	AP07K.11	81.95	78.77	85.87	78.43	89.92	91.58	91.44	91.47	90.91	90.56	90.54	90.59	90.65	90.70	90.16	90.29	90.23	90.29	93.27	99.67	99.80	99.80	99.80	99.93	100.0	100.0	81.62	82.38	82.96	83.32	
	AP07K.225	82.21	78.77	86.00	78.43	90.06	91.69	91.56	91.47	91.03	90.56	90.54	90.72	90.65	90.83	90.30	90.43	90.23	90.43	93.36	99.67	99.80	99.80	99.80	99.93	100.0	100.0	81.73	82.57	83.19	83.36	
	<i>P. thermos.</i>	79.07	79.34	79.75	80.42	81.99	81.46	81.40	81.38	81.00	81.74	80.68	81.09	81.63	81.70	81.77	81.73	81.90	81.05	81.53	81.52	81.73	81.66	81.73	81.62	81.73	100.0	88.76	89.26	89.55		
	<i>R. hungatei</i>	79.37	79.00	79.56	79.77	82.46	82.34	82.14	81.99	81.27	81.53	80.46	80.96	81.56	81.78	82.52	82.52	82.34	82.59	81.75	82.36	82.29	82.57	82.50	82.57	82.38	82.57	88.76	100.0	89.47	89.48	
	<i>H. aldrichii</i>	81.42	80.73	82.02	81.39	83.28	84.05	83.98	83.98	83.92	84.05	81.68	82.15	83.03	83.21	84.10	84.03	83.82	84.10	83.45	83.12	83.01	83.05	83.25	83.19	82.96	83.19	89.26	89.47	100.0	97.87	
	<i>H. cellulol.</i>	81.67	81.06	82.14	81.87	83.31	84.03	83.96	84.04	84.10	84.55	82.10	82.37	83.46	83.45	84.21	84.28	84.25	84.35	83.34	83.29	83.37	83.22	83.43	83.36	83.32	83.36	89.55	89.48	97.87	100.0	







Highlights

- Known: Ct85-amplicon abundancies differ between subject groups in IBS studies
- Specific primers designed for six sequence types in the Ct85 16S rRNA gene cluster
- Sequence types in the Ct85-cluster tentatively represent species, genera, families
- Ct85-cluster sequence types found in the guts of humans, pets, production animals
- Sequence type composition of Ct85 differs across human and animal populations

Conflicts of interest

The authors declare no conflicts of interest.